sciendo

NS global economics

ECONOMICS
Volume 10, No. 2, 2022

www.economicsrs.com

CUSTOMER CHURN PREDICTION MODEL: A CASE OF THE TELECOMMUNICATION MARKET

Yana Fareniuk [1], Tetiana Zatonatska [1], Oleksandr Dluhopolskyi [2,3], Oksana Kovalenko [1]

Received 22. 08. 2022. | Sent to review 01. 09. 2022. | Accepted 04. 11. 2022.

Original Article

[1] Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

[2] WSEI University, Lublin, Poland

[3] West Ukrainian National University, Ternopil, Ukraine

Corresponding Author:
Oleksandr Dluhopolskyi
Email: dlugopolsky77@gmail.com

JEL Classification:
C59, D11, M31

Doi: 10.2478/eoik-2022-0021

UDK: 621.39:[658.8:339.13.017

ABSTRACT

The telecommunications market is well developed but is characterized by oversaturation and high levels of competition. Based on this, the urgent problem is to retain customers and predict the outflow of customer base by switching subscribers to the services of competitors. Data Science technologies and data mining methodology create significant opportunities for companies that implement data analysis and modeling for development of customer churn prediction models. The research goals are to compare different approaches and methods for customer churn prediction and construct different Data Science models to classify customers according to the probability of their churn from the company's client base and predict potential customers who could stop to use the company's services. On the example of one of the leading Ukrainian telecommunication companies, the article presents the results of different classification models, such as C5.0, KNN, Neural Net, Ensemble, Random Tree, Neural Net Ensemble, etc. All models are prepared in IBM SPSS Modeler and have a high level of quality (the overall accuracy and AUC ROC are more than 90%). So, the research proves the possibility and feasibility of using models in the further classification of customers to predict customer loyalty to the company and minimize consumer's churn. The key factors influencing the customer churn are identified and form a basis for future prediction of customer outflow and optimization of company's services. Implementation of customer churn prediction models will help to maintain customer loyalty, reduce customer outflow and increase business results

Keywords: *marketing, classify customers, telecommunications market, machine learning, prediction, Data Science models.*

1. INTRODUCTION

The volumes of mobile communication services and telecom enterprises are growing rapidly due to significant technological advances. Telecommunications has become a lucrative but aggressively competitive market. Currently, this market tends to saturate as supply exceeds demand (Fang, 2021), and the main source of business is achieved through the switching between operators. As the number of mobile operators increases, consumers can switch from one telecommunication operator to another if they are not satisfied with the service. Competition between operators for current consumers is becoming increasingly intense (Ding, 2022).

Clients are a highly important asset in any business. To maintain their market position, companies must satisfy requirements of their customers. Customer retention strategy plays a crucial role in attracting and retaining customers, a company's business growth and means that the customers are satisfied, and they don't stop using company's services (Saanchay et al., 2022). In a competitive market for mobile telecommunications, the customers want competitive pricing and high-quality services. Without hesitation, the customer will change his telecommunications service provider if he / she does not find what he / she is looking for (Kolli et al., 2020).

The economic activity of the enterprise is related to innovative management, which largely determines the success of its functioning (Rakhmatullina et al., 2022). In today's business environment, which is characterized by a high level of uncertainty, understanding consumer behavior is an important part of the process of strategic planning and business activities. This is a key factor in achieving market leadership, so it is important to focus business processes on the customer. By implementing data-driven marketing, companies can find the right customers, retain, and develop them, and ensure business growth with a sufficient profitability level (Grandhi et al., 2021).

Customer outflows (churn) and its forecasting are one of the most challenging and overriding problems for the telecommunications industry that affects revenue and the company's growth. According to a recent Gartner Tech Marketing survey, 91% of respondents rate customer churn as one of their top concerns. However, only 43% have invested in additional resources to this problem's solving (De et al., 2021).

The key to surviving in this competitive industry is to know customers better and retain every valuable customer (Qu, 2022). Many valuable customers can be lost to competitors in the telecommunications market, leading to risk of losing profits (Fang, 2021). The marketing literature proves that acquiring a new customer is 5-10 times more expensive than retaining an existing one. Thus, effective management of customer churn and understanding the reasons for customer outflows have become a quite crucial task for mobile operators (Thakkar et al., 2022; Zhang et al., 2022). Customer relations and customer knowledge have positive influence on company's performance and service quality of the telecom enterprises (Abd-Elrahman et al., 2022).

The main classification problem in the telecommunication industry is related to the task of subscribers' churn prediction and churn prevention (Kuznietsova et al., 2022). Companies need to focus on improving their information and communication technologies, experience and offered rewards for better customer retention (Abarca Sánchez et al., 2022). The company can also concentrate its marketing activities on encouraging customers to switch to the latest technologies, which can lead to cheaper costs (Cambier et al., 2022). Improving the efficiency of customer's interaction with companies and minimizing the outflow of customers are relevant and important, so there is increasing role of modern data mining methods, models, tools, algorithms, and machine learning technologies, which generate additional opportunities in this area (Kolomiiets et al., 2021; Havrylovych et al., 2019). Customer retention depends considerably on data analytics and predictive modeling to support decision making process (De Caigny et al., 2021).

The emergence of new technologies generates technological changes in the telecom market. Companies are dealing with new opportunities and challenges in their development (Li, 2022). Information technologies are steadily entering the business environment, there is a rapid development of predictive analytics, data analysis, telecommunications, as the economic value of information is related to business success and emerging business management opportunities (Agafonova et al., 2022). With advances and the explosive growth in data mining technologies and the availability of vast amounts of customer data, companies have begun to use powerful data science tools in precision marketing to dive deeper into the nature and habits of their customers. This strategy allows them to personalize marketing and advertising information for their consumers needs through appropriate recommendation development (Qu, 2022; Gu, 2022). Companies can also track changes in customer behavior so that they can make strategic decisions in advance to retain valuable customers.

It is extremely important for companies to maintain the portfolios of current customers together with attracting new customers. In addition, a useful method of retaining existing customers will make the company more profitable by reducing the total cost of marketing and advertising to compensate for attracting new customers. Recent research suggests that customer retention is an important aspect of improving operational efficiency, and that various service sectors need to focus

better on retaining their current clients (Todevski et al., 2021).
Using precision marketing and machine learning technologies for data analysis, it is possible to develop a more effective marketing strategy for retaining existing consumers of telecom operators by accurately predicting its consumer churn behavior, analyze current problems and identify appropriate recommendations and suggestions to optimize the activities of the telecommunications company.

2. LITERATURE REVIEW

Recently, the marketing concern about retaining their customers has grown significantly, as it is a crucial marketing issue. Recent marketing data indicate that attracting (acquisition) new customers or subscribers is much more expensive and complicated than retaining current ones in the competitive business environment (Kumar et al., 2020). Thus, the loss of a client will negatively affect the business growth and its profitability (Al-Shatnwai et al., 2020).
At the current stage, as market saturation increases, all companies should focus on identifying customers by their potential outflow as part of their business strategy. Early and accurately prediction of customer behavior, anticipating which consumers intend to churn the company's services and understanding the reasons for this intention play an important role in marketing, which aims to maintain loyal customers (Gopal et al., 2021; Vezzoli et al. 2020). It can be achieved by extracting (mining) information that affects the outflow from customer data. As a result, a huge number of scientific papers on this topic have been created.
The attention of both researchers and marketing practitioners is increasingly focused on forecasting the outflow of customers, which remains an important part of the process of maintaining the customer base and is a popular approach for solving the problem of customer churn (Fridrich, 2020). The marketing process, accompanied by the rapid accumulation of large amounts of data and information in the telecommunications industry and the growing maturity of Data Mining, contributes to the wider application and development of customer churn models to predict customer behavior by Data Mining techniques. In this way, the telecommunications company can effectively predict the outflow of customers to ensure that the outflow of customers is avoided by forming effective marketing measures (Tianyuan et al., 2021) and implementing successful marketing strategy (Xiahou et al., 2022).
In the telecommunications sector, a huge amount of data is generated daily due to a large customer base, and innovation capabilities play an important role in marketing (Aljanabi, 2022). Many authors in their previous research have presented a variety of models for predicting customer outflow (churn) using data mining and machine learning technologies. Tianyuan et al. (2021) proved that the most widely used data mining methods for predicting customer outflows are logistic regression, decision tree, support vector machine (SVM).
In the last decade, many scientists present the application results of various machine learning methods and algorithms for classification (Bandam et al., 2022) and prediction of churn behavior of the most valuable part of the current clients (Günesen et al., 2021), searching for more efficient approaches of customer churn prediction. Among the used methods and models for solving the tasks of data processing for telecommunication industry are logistic regression, decision tree and random forest models for churn prediction (Kiguchi et al., 2022; Vezzoli et al., 2020; Kuznietsova et al., 2022), K-means, SVM (Sánchez et al., 2022), the combination of k-means customer segmentation and SVM prediction (Xiahou et al., 2022), the multi-level classification using SVM in the SLS-SVM algorithm (Huang, 2022), the Naïve Bayes for prediction of loyal or disloyal customers and their behavior (Jayadi et al., 2020; Rabiul Alam et al., 2021), an ECHAID (exhaustive Chi-square automatic interaction detector) classification tree for consumers segmentation (Kelley et al., 2022) and so on.
Zhang et al. (2022) used Fisher discriminant equations and logistic regression analysis to develop

a customer churn prediction model through customer segmentation on the telecommunication market. Kuznietsova et al. (2018) constructed gradient boosting for estimating the probability of the consumer's outflow, which shows high quality performance. Zhu et al. (2020) proposed a trajectory-based deep sequential method TR-LSTM and utilized the long short-term memory neural network (LSTM) to conduct sequential modeling. Thakkar's et al. (2022) study proposes AdaBoostWithCost algorithm, which improves the discrete AdaBoost algorithm for churn prediction on telecom market. Research by Al-Shatnwai et al. (2020) proposes one of the most powerful machine learning classifiers XGBoost as a customer retention model.

Also, recently, artificial intelligence methods have become more widespread in predicting the customer churn for telecommunications, in particular, artificial neural networks (ANN) are recognized as a very effective method of forecasting and classification of churners and non-churners (Saanchay et al., 2022). The research by Kumar et al. (2020) proposes a rule-based customer churn prediction model using ANN. The application of the proposed framework will help enterprises in development and implementation of intelligent decision support system. Cacciarelli's et. al. (2022) study compares the results of classification by the combination of sampling techniques and decision trees, ensemble methods and ANN.

In order to solve the existing marketing tasks related to customer management, minimization of customer churn and optimization of marketing costs, Park et al. (2022) and Mo et al. (2022) investigate the relationship between different factors influencing risk of customer outflow by deep learning technologies. Dadfarnia et al. (2020) present the application of Deep Neural Network, which has significantly better accuracy vs other methods. Scientists proposed several classification models and boosting methods to control customer churn rate.

Hemalatha et al. (2020) proposed an outflow prediction model that uses classification and clustering approaches to identify customer churn and determine the factors that cause customer outflows in the telecommunications using Adaptive Logitboost. By identifying key influencing factors of consumer's churn, companies can increase productivity by developing appropriate customer recruitment advertising campaigns based on the association between behavioral patterns. In this way, companies increase the quality of marketing campaigns.

Priyanga et al. (2022) suggested the Hierarchical Flexi-Ensemble Clustering model. Deng et al. (2021) create a model for churn prediction by ensemble learning algorithms such as Lightgbm, Catboost, Random Forest to improve the accuracy of forecasting and achieve cost savings. Rezaeian et al. (2021) proposed ensemble models with decision trees to predict customer churn. The research by De et al. (2021) shows that although hybrid and ensemble methods have been widely successful and have improved model performance, there is a need for clearly defined recommendations for appropriate model evaluation measures.

Kuznietsova et al. (2022) identified neural networks, SVM as promising methods to solve problem of customers classification by their preferences and services, as well as the tendency to churn. The dynamic approach based on dynamic models of survival theory for churn time forecasting is also proposed. Havrylovych et al. (2019) suggested an approach that provides information not only about the probabilities, but also about the time when there may be an outflow of customers. To achieve this goal, two algorithms have been developed based on the use of survival functions and prediction of the outflow time. It also offered a step-by-step process of churn in the decision support system to detect outflows and timely identify the most dangerous groups of customers who are at risk of outflow.

The paper by Kuznietsova (2017) shows the possibility of using information technology to analyze the database of subscribers of the telecom company to predict their future behavior. The task is an actual not only in terms of predicting the fact of changing the subscriber of the telecommunications operator and refusing to use the services, but also the moment when the customer just started thinking about it. The article solves two problems: the task of classification (the problem of predicting the possible customer churn) and the problem of predicting the time when it may occur.

A generalized linear model was constructed and showed acceptable predictive properties based on the GINI index, but lower in comparison with logistic regression, neural networks and gradient boosting. Knowing the risk period will be useful for the business in terms of preventing the customer churn by developing personalized offers and providing additional incentives for current clients.

The conclusions of the research by Jamjoom (2021) prove that data mining methods can be highly successful in finding hidden insights and highlighting customer's information. Besides, it is found that each model works effectively with a different training dataset. Customer outflow forecasting model using machine learning for classification has been developed mainly based on a single time piece of data, but Gattermann-Itschert et al. (2021) proposed the approach, which involves model's construction based on multiple time pieces of data. It should be noted that all findings and conclusions from modeling and Data Science implementation cannot be taken absolutely for all companies on the market, as the results are a consequence of many factors and conditions that are established at each time, which requires an adapted approach in each case (Chernyak & Fareniuk, 2020).

A deep understanding of customer's needs and effective analysis of their behavior are tools that contribute to the success of every business in the market. The ability to retain customers involves effectively studying current and potential customers, making more effective decisions and improving business processes. With a thorough understanding of customers and their behavior as consumers, a business can maintain customer loyalty, predict how customers will respond to a company's marketing strategy, improve business strategy, and increase profits. The study by Rabiul Alam et al. (2021) examined the purchasing habits of telecommunications customers to help improve their position in the telecom market.

The development of technology has provided a highly competitive marketing environment, where the analysis of consumer behavior has become vital. At the same time, the number of financial resources spent on customer retention is much less than on attracting new customers (Goy et al., 2020). With the development of intellectual potential, an effective customer relationship management (CRM) system based on data mining technologies can generate more economic benefits to the company by improving the level of management and business and marketing decisions. Optimal customer retention strategy helps companies to adjust marketing strategies and increase efficiency (Sun et al., 2022; Radukic et al., 2019). Abarca Sánchez et al. (2022) aim to establish the relationship between loyalty and retention in telecom companies through an exploratory factor analysis and a multiple linear regression modeling.

Du et al. (2022) proposed a quantitative Kano model to analyze consumer personal demand and achieve a high validity of marketing decisions. Authors recommend providing a personalized proposal for each heterogeneous customer group. Customers are classified as groups, which prefer price, brand, or service as a priority. As a result, it is relevant to adapt marketing strategies and develop appropriate solutions for each group of consumers. Determining various consumer types and implementing differentiated marketing strategies can help businesses increase profitability and customer loyalty. Recency, frequency, and monetary model is an important method of data mining that has significant practical value in CRM and marketing (Hu et al., 2022). Wu et al. (2021) proposed framework with customer-value-weighted machine learning models, which give companies useful insights to more effective development of marketing strategies for various consumer's groups.

In precision marketing, the allocation of limited resources to different target groups of customers is difficult. Zhang et al. (2022) presented a framework using a distance-based algorithm, K-nearest neighbors, and SVM to capture customer preferences for promotion channels. A resource optimization model using classification results was developed.

Belbahri et al. (2021) proposed a Qini-based uplift model based on logistic regressions, which are used to identify customers who are more likely to respond positively to targeted marketing activities to retain them, i.e., to avoid unnecessary costs for customers who are more likely to switch on competitors. The Qini coefficient measures the model performance.

Taking into account the increasing of customer behavior importance for businesses, telecom companies are focusing not only on increasing customer profitability, increasing market share, but also on working with highly loyal customers and customers who can switch to competitors (churn). There is the need to transform marketing tools for customer experience management in modern conditions (Syaglova et al., 2022). The development of big data concepts has shaped a new wave of customer relationship management strategies. Big data analysis provides an opportunity to investigate customer's behavior, understand their preferences in order to develop relevant marketing plans (Wassouf et al., 2020).

Big data analytics is becoming a strategic tool for achieving business efficiency and implementing data-based marketing. Ram et al. (2022) argue that the introduction of data analysis helps to increase customer lifetime value, respond to market fluctuations and changes, improve the process and quality of management. Nowadays, there is a significant need to implement advanced data-driven decisions based on Data Science technologies, as simple analytics will not cover the business demand (Fedirko et al., 2021; Mašić et al., 2018). In order to solve the low stability and efficiency, high complexity of marketing decisions, a marketing adaptive decision-making algorithm based on big data analysis is proposed by Lv (2022).

Using intelligence results from customer churn prediction models, there is the significant potential to generate additional revenue by improving customer retention strategy (Günesen et al., 2021). Churn prediction model will help telecom enterprises to effectively forecast the opportunity of customer's churn, take appropriate targeted measures and make better marketing decisions to avoid the outflow of customers and, as a result, increase their profits.

Despite the significant amount of scientific research, many issues of customer churn modeling and client's classification remain unclear the effectiveness of different data mining methods for improvement of marketing activities of churn management. The uncertainty and instability of economic dynamics, desire for continuous development have challenged all businesses to implement Data Science into marketing system of decision-making.

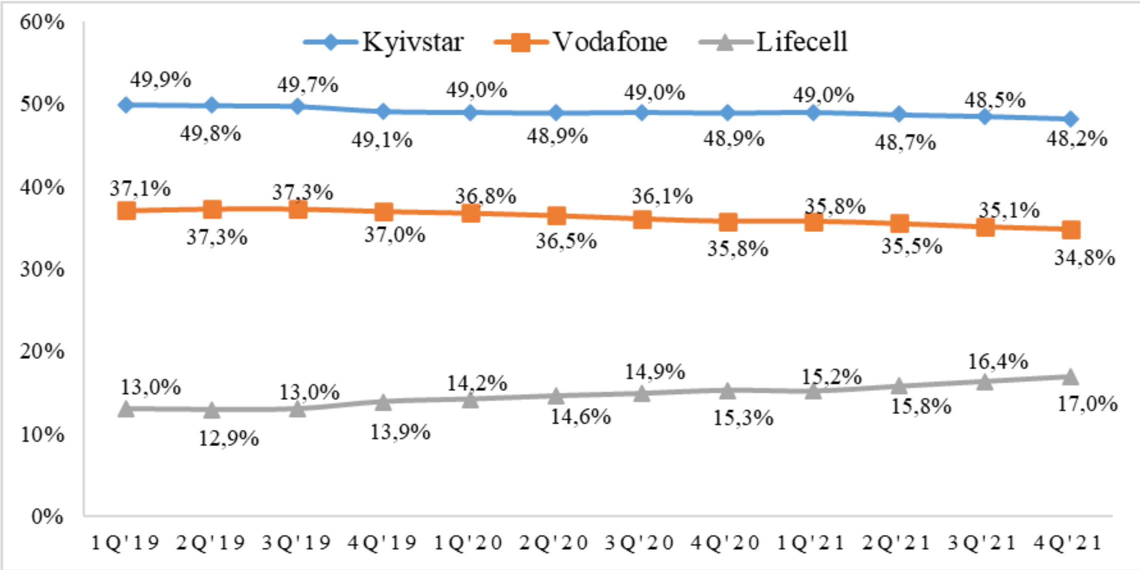
3. AIM OF THE RESEARCH

The purposes of this study are to compare different approaches and methods for customer churn prediction and build various Data Science models to classify customers according to the probability of their churn from the company's customer base and predict customers who could potentially refuse the company's services.

4. METHODS

Over the last years, the share of the telecommunications industry in Ukraine has grown significantly, but telecommunication companies operate in conditions of fierce competition. Major players in the Ukrainian market are Kyivstar (48%), Vodafone (35%), Lifecell (17%) (Figure 1).

Figure 1. Dynamic of the Kyivstar, Vodafone and Lifecell market share, 2019-2021.



Source: (Zatonatska et al., 2022).

Mobile penetration in Ukraine is approaching 150%. This situation is explained by the tendency of subscribers to use the services of several operators, for example, to have two SIM cards of different operators at the same time (Khrustalova et al., 2019). This, in turn, means that every year it becomes more and more difficult for operators to attract new subscribers. New connections are more the result of stimulating subscribers who are more sensitive to price offers. Attracting a new subscriber is an additional cost for the operator. Thus, the task of retaining an existing subscriber has certain economic grounds.

The driving force behind the development of the telecommunications industry is scientific progress, as this industry depends entirely on the technical support used by a particular company. It is technical support that determines the quality of services provided by a company. Leading companies in this industry conduct market research and constantly improve their services in terms of quality, as well as make them cheaper. The key to success is the stability and high quality of communication, their low cost, and the flexibility and versatility of services.

The telecommunications company has accumulated statistical information about subscribers. Considering the problem of retaining an existing subscriber of a telecommunications company, the company's management has a task to develop classification models that would reduce the outflow of subscribers, i.e., maximize customer loyalty of the telecommunications company. Business goals for this research include correct construction of the classification model; reduction of customer churn by 10%; growth of consumer loyalty; creation of new competitive advantages; reducing the cost of retaining existing customers by creating an optimal personalized offer for each group (segment) of consumers, which has risk of churn.

In order to maintain market leadership in conditions of difficult economic situation, changing of consumers behavior and high level of competition, a set of key strategic marketing goals includes gaining new customers, retaining current customers, introducing new technologies and services, continuing to develop the brand and improve its market position, paying attention to customer wishes and providing maximum support; offer exactly what the customer needs and anticipate his wishes. However, high competition and market oversaturation, offering cheaper services by major competitors are main threats for stable company's growth.

One of the main tasks of churn management is to construct an effective model for churn prediction that can determine clients who are most likely to outflow. The customer churn determinants can be classified into several groups: service quality (i.e., tariff, customer service value added, etc.), brand image (i.e., innovative, competitive aspects of the telecom company) and consumer's behavior (level of services consumption by different subproducts, by time intervals, etc.). The key idea is to investigate customer's profile using different data sources including patterns of calls, tariff information, costs, payment, customer service calls, demographic variables and then forecast the probability that consumers will churn based on their features (Kolli et al., 2020).

To achieve the research goals, it is necessary to implement the relevant model, using system approach with CRISP-DM methodology, which describes the process through 6 main stages: business understanding, data understanding, data preparation, modeling, evaluation and deployment.

Available information and data from one of the main telecommunication companies in Ukraine (among TOP-3) include customer database of this company, which contains impersonal customer data about their activities (calls, SMS, internet, call duration, etc.), most frequently used services, price category of the favorite tariff, interests of customers, about the devices they use, location, consumer's demographics. The process of dataset preparation involved work with the accumulated data, in particular analyzing of outliers and errors in the data, as well as removing the unnecessary data that do not have significant fluctuations between different consumers (for example, device) or do not have a significant impact on consumer's outflow (for example, location as all operators on the telecommunication market are national). As a result, the received data have 4492 rows of data with 11 main variables: age, average monthly expenses, average duration of conversations, calls during the day for a month, calls in the evening for a month, calls at night for month, calls to landlines per month, calls abroad per month, share of calls to other operators per month, number of SMS per month, outflow as a Boolean variable. The description of each variable is presented in Table 1.

The database contains information about customer tariff plans and their corresponding other characteristics, as it contains information about the number of calls at different times of the day, calls abroad, calls in the network, etc. By comparing which, it will be possible to determine the most common tariff plan and compare it with the relevant target group. The scientific hypothesis of the research is that collected information about the activity of the company's customers regarding the use of telecommunication services is useful for construction an effective consumer churn model, which can be implemented by telecom operators to optimize their marketing activities.

Table 1. Description of variables in dataset of telecommunication company.

Variable	Description	Type of data
Age	Age of client	Integer
Average monthly expenses	Average amount of the subscriber's expenses for mobile communication and mobile Internet (tariff plan + additional expenses)	Continuous
Average duration of conversations	Average number of minutes for outgoing calls spent by the subscriber per month	Continuous
Calls during the day for a month	Number of calls per month in the morning and afternoon	Integer
Calls in the evening for a month	Number of calls per month in the evening	Integer
Calls at night for a month	Number of calls per month at night	Integer
Calls with other operators per month	Number of calls per month with other operators of Ukraine	Integer
Calls abroad per month	Number of calls per month with other countries	Integer
Calls to landlines per month	The share of calls to landlines	Continuous
SMS per month	Number of SMS in subscriber's consumption	Integer
Outflow	An indicator of subscriber outflow (churn) from the consumer base. 1 – a subscriber, who stopped using the services of a telecommunications company (ex-client); 0 – current client. Total amount of consumers – 4492 and 444 are described as churners.	Boolean (1/0)

Source: constructed by authors based on internal database of telecommunication company.

The marketing task is the classification to determine the most attractive tariff plan for customers. Success criteria of classification model construction and development are: increase the number of new customers by 5%, reduce the outflow and increase the level of NPS by 10%. The outflow level of more than 15% is quite critical for the company, so it is relevant to focus marketing activities on maintaining consumers. Identifying the most loyal customer groups and proposing them the best tariff plan will help to meet the needs of the selected group and increase marketing efficiency. IBM SPSS Modeler was selected as the appropriate software for classification model's development and submission.

Data preparation also involves the creation of additional fields in the database. Since the age of respondents is from 19 to 70 years, we divide the age of subscribers into the following 5 categories (variable Age Category): 19-24 years, 25-34 years, 35-50 years, 51-60 years, 60-70 years. We also create variables Average monthly costs (normal) – normalized values of the average cost of the call, and Average duration of conversations (normal) – normalized values of the average duration of the call.

Customers can be segmented by time, frequency, and monetary aspects. Time includes duration of calls and internet connections during a certain period. Frequency covers that consumer uses company's services frequently during a certain period. The monetary aspect contains expenses level during a certain period. These aspects for segmentation also were used in the new TFM (Time-frequency-monetary) approach, proposed by Wassouf et al. (2020). By analyzing the signs of loyalty at each level, it is possible to offer consumers the most relevant offers and services.

When choosing a modeling method, it should first be noted that the purpose of modeling is to predict customer loyalty to mobile operators. Analyzing the available data types and the list of requirements for modeling identified the following promising methods of data mining: Auto Classifier (including logistic regression, decision trees, XGBoost), C5.0, KNN, Neural Network, Ensemble, Random Tree, Ensemble Neural Network and so on. Such Data Mining methods was selected as the most relevant due to their high performance and usefulness for customer outflow modeling as we mentioned in literature review. Data mining technologies were used to explore clients and achieve knowledge extraction in forecasting customer outflows.

For evaluating churn prediction models, the following parameters were used: accuracy, precision, first and second-types errors, F-measure, and area under the receiver operating characteristics curve (AUC ROC), which is the most effective criterion and combines the possibility to receive high accuracy, sensitivity to real situation and minimize the specificity of the model (Kuznietsova et al., 2018; Hemalatha et al., 2020).

A model will be considered successful if at least 90% of customers are classified correctly. In addition, the effectiveness of the models will be defined as the ratio of income that the firm will have in the case of the implementation of a model of customer classification to the income that the firm will have in the current case without model implementation.

customer churn modeling Results

We started model's construction from the AutoClassifier node in IBM SPSS Modeler. This node allows you to evaluate and compare several different modeling methods, allowing you to test different approaches in one run of the modeling process. For example, instead of choosing one of the neural network options (fast, dynamic or truncated), you can try all of them. This node generates and explores a set of models based on all possible combinations of options, ranks candidate models based on a given indicator (the specified criteria) and saves the best models for further analysis.

You can choose the appropriate option from the three automatic modeling nodes based on your analysis needs. The autoclassification node creates and compares several different models for binary output (yes or no, the client will churn or stay with company, etc.), allowing you to choose the best approach for customer churn analysis and prediction. In this research, we selected all possible models (C5, logistic regression, Bayesian network, KNN algorithm, LSVM, random forest, SVM, XGBoost tree, CHAID, Quest, etc.).

As a target field we selected *Outflow* variable and as an input fields - average duration of conversations, daytime calls per month, evening calls per month, night calls per month, landline calls per month, calls abroad per month, share of calls to other operators per month, number of SMS per month. As criteria for model deviation, we require that the area under the curve (AUC ROC) be at least 0.8 (80%). We also use income criteria when setting up an AutoClassifier node, including cost criteria at level 5 and profit at level 10. Results of TOP-10 models from AutoClassifier node presented in Table 2. So, after analyzing the results obtained using the AutoClassifier node, we identified the 5 best models: C5, XGBoost Tree, Quest, CHAID, Random Trees, Neural Network, in which the overall accuracy exceeds 97% and the area under the curve is greater than 0.995.

Next, let's analyze the importance of predictors. The most predictors for all constructed models are equivalent, but the most important predictor is average duration of conversations.

In C5.0 model the most important predictors that indicate a potential subscriber's outflow from the company's network are average call duration and share of calls to other operators. The remaining 8 proposed attributes were eliminated during the operation of the C5.0 model (Figure 2).

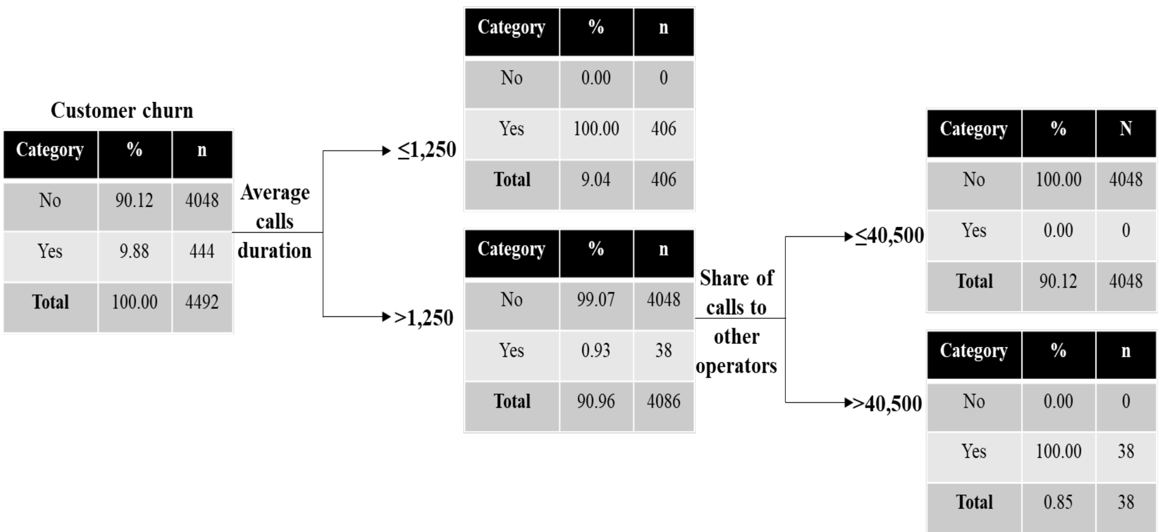
Table 2. The results of comparing the quality of TOP-10 AutoClassifier models.

Model	Maximum win	Common accuracy	Number of correctly predicted churners	AUC ROC
C5	2195	100	444	1
XGBoost Tree	2195	100	444	1
Quest	2195	99.911	444	1
CHAID	2020	99.154	443	0.999
Random Forest	1890.75	97.996	443	0.996
Neural Net	1850	98.175	404	0.995
KNN	1420.385	96.549	290	0.986
SVM	1275	95.347	271	0.980
Logistic regression	1320.0	95.481	281	0.960
LSVM	1140.0	94.724	243	0.959

Source: authors calculation in IBM SPSS Modeler based on internal database of telecommunication company.

After analyzing the results, we obtained that the overall accuracy of the classification by C5.0 is 100%, AUC is 1, which means an excellent result of data classification. The overall accuracy of the constructed model by KNN is 96.13%, the area under the ROC-curve – 0.983, which indicates a fairly accurate classification of customer loyalty (Table 3).

Figure 2. C5.0 decision tree, where the first distribution is based on the most important attribute – Average duration of conversations.



Source: constructed by authors based on results of model development.

Table 3. The results of quality assessment of the C5.0 model and KNN model.

C5.0		
	No	**Yes**
No	4 048	0
Yes	0	444

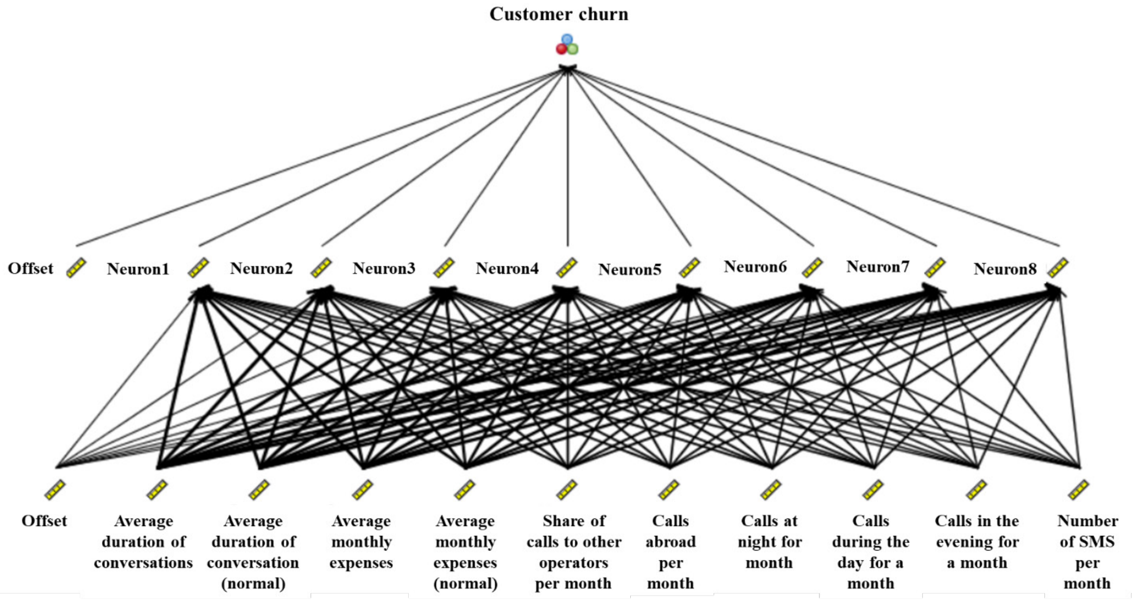
KNN		
	No	**Yes**
No	4 028	20
Yes	154	290

Source: constructed by authors based on results of model development.

The next one model is Neural Network. In a typical neural network, there are three parts: the input layer neurons represent the input fields, there are one or more hidden layers, and there is an output layer containing one or more neurons representing the target fields. Each connection between neurons is assigned a force of influence or weight. The input data enters the first layer, then the values are propagated by the layer from each neuron of this layer to each neuron of the next layer. The result is obtained from the output layer.

In this research, a multilayer perceptron was constructed for the target metric outflow and the model was chosen according to the rule "It is impossible to reduce the error". The general view of the constructed Neural Network model is presented in Figure 3. Table 4 presents the results of model quality assessment, where the overall accuracy of the model is 98.2%, AUC is 0.995.

Figure 3. General view of the constructed Neural Network model.



Source: constructed by authors in IBM SPSS Modeler.

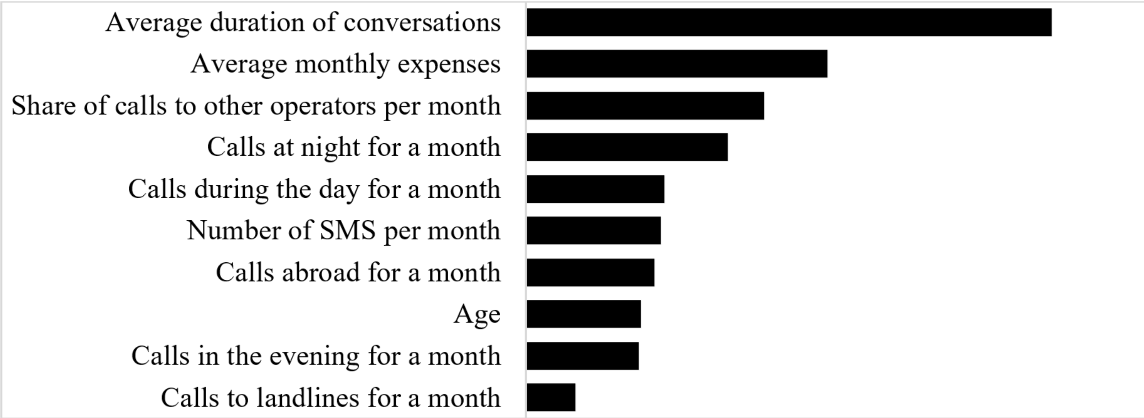
Table 4. Table of losses of erroneous classification of the Neural Net model.

	No	Yes
No	99.0% (4 006)	1.0% (42)
Yes	9.0% (40)	91.0% (404)

Source: constructed by authors based on results of model development.

Analyzing the results of the importance of predictors, we obtained that the most important predictor is the average duration of conversations, average monthly costs, and a little less important – the share of calls to other operators per month (Figure 4).

Figure 4. The importance of predictors in the Neural Net model.



Source: constructed by authors based on results of model development.

The Ensemble node combines several model casts, which allows you to get more accurate predictions than any of these models alone. By combining predictions from several models, you can avoid limitations in individual models, which will lead to higher overall accuracy. In this case, let's create an ensemble of C5.0, KNN and Neural Net models and compared the results.
Thus, the accuracy of the model is 98.84% (Table 5), which is higher than when building a separate Neural Net (97.5%) or KNN (96.13%) models, but slightly lower than when building a separate model C5.0 (100%).

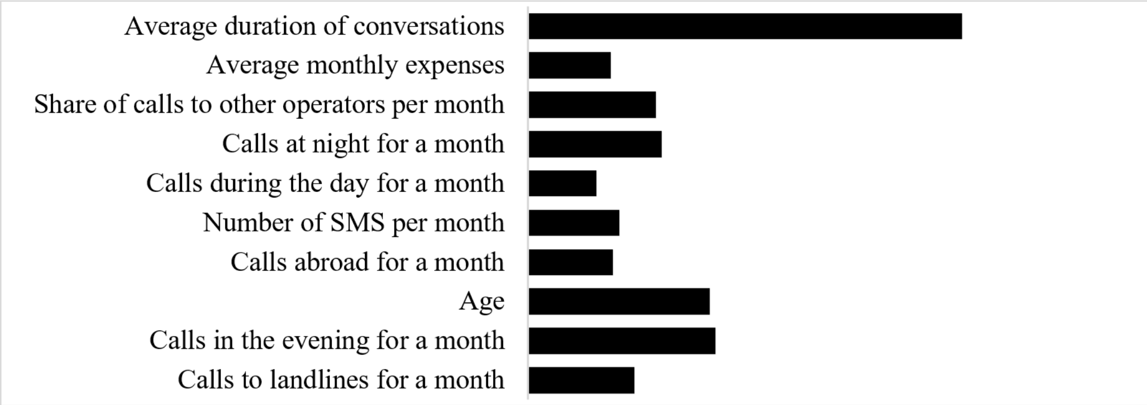
Table 5. The results of quality assessment of the ensemble of models.

	No	**Yes**
No	4 041	7
Yes	45	399

Source: constructed by authors based on results of model development.

The next one node is Random Tree. Decision tree models are used to create classification systems that predict or classify future observations based on a set of decision rules. This approach has several advantages. First, the reasoning process underlying the model is seen when looking at the tree. Second, the process automatically includes in its rules only those attributes that matter when making a decision. Attributes that do not contribute to the accuracy of the tree are ignored. Molds of the decision tree model can be converted into IF-THEN rules (a set of rules), which in many cases provides information in a more understandable form.
We analyzed the results of using the method of random trees to classify customer loyalty and potential outflow from the network. Thus, the estimate of the accuracy of the model is 96.9%, the estimate of the proportion of erroneous classifications is 0,029. From Figure 5, we can see that average duration of conversations is the main predictor.

Figure 5. The importance of predictors in the Random Tree model.



Source: constructed by authors based on results of model development.

Main rules of decisions for ouflow variable, for which the most common category is *No*, and the accuracy of the rules, the accuracy of the ensemble and the level of interest = 1 are presented below (Table 6):

1. (Average monthly expenses (normal) <= 0.87) & (Share of calls to other operators per month> 23.0) & (Average call duration (normal)> -0.98);
2. (Share of calls to other operators per month <= 23.0) & (Average call duration (normal) > -0.98);
3. (Calls per day per month ≤ 104.0) & (Average call duration > 1.3);
4. (Calls per day per month ≤ 104.0) & (Average call duration (normal) > -0.98) & (Age > 23.0);
5. (Share of calls to other operators per month ≤ 23.0) & (Average monthly expenses > 119.25) & (Age ≤ 23.0).

Table 6. The results of quality assessment of the Random Tree model.

	No	**Yes**
No	3 959	89
Yes	1	443

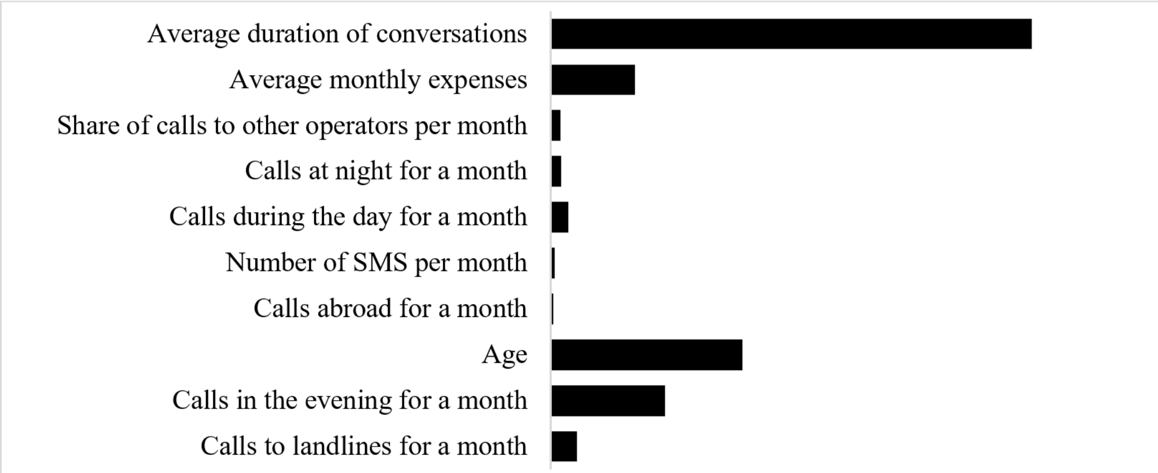
Source: constructed by authors based on results of model development.

Thus, the overall accuracy of the model is 98%, AUC – 0.996, Ginny – 0.993, which indicates a fairly high level of classification of customers by potentially possible outflow from the company's network.

Let's try to increase accuracy by hybridizing neural networks with Ensemble Neural Net. To do this, we chose the boost method. In this way, we created an ensemble using boosting, which generates a sequence of models to obtain more accurate predictions.

As a result, the accuracy of the ensemble and the reference model reaches 100%, the naive model – 90.1%. The most important predictor is the average duration of conversations (Fig. 6).

Figure 6. The importance of predictors in the Ensemble Neural Net model.



Source: constructed by authors based on results of model development.

According to the results of the constructed model, we proved the extremely high accuracy, reaching 100%, the AUC also has the maximum possible value of 1, respectively.

6. DISCUSSION

Comparison of accuracy of models and evaluation of results. Since the lowest allowable quality of models for research purposes was defined as 90%, the results of the obtained models (the quality of which exceeds 95%) fully meet the objectives (Table 7).
As a result of the research, we confirmed the hypothesis that the average duration of conversations has the greatest impact on the client's churn from the company's network, so the target indicator of improving tariff plans will be the cost of calls and the quality of communication.

Table 7. The summarized results of the accuracy of the constructed models.

Methods	Accuracy	AUC	GINI
C5.0	100%	1	1
KNN	96,13%	0,983	0,966
Neural Net	98,17%	0,995	0,999
Ensemble	98,84%	1	0,999
Random Tree	98%	0,996	0,993
Ensemble Neural Net	100%	1	1

Source: constructed by authors based on results of development of different models

So, comparing the quality of the constructed models, it should be noted that the most accurate data are obtained using an ensemble of neural networks (total accuracy 100%), C5.0 (total accuracy 100%) and an ensemble consisting of a combination of three models (total accuracy 98.84%).

7. CONCLUSIONS

Customer outflow management is a vital task of marketing management of a telecommunications company, which is developing in conditions of oversaturation of the market and fierce competition. To ensure the effective functioning of companies, it is necessary to implement approaches, methods and models of data mining and Data Science.

Implementation of Data Mining models and each stage of the process based on the CRISP-DM methodology provided a qualitative end result of the study. At the start of the research, the goals and the peculiarities of the churn prediction, the main issues, the criteria for the success of the model, possible threats were identified. At the modeling stage, the main models (C5.0, KNN, Neural Net, Ensemble, Random Tree, Neural Net Ensemble) were implemented with the help of IBM SPSS Modeler. Estimation of results accuracy reached over 95%, which indicates the possibility and feasibility of using models in the further classification of customers to determine customer loyalty to the company and minimize consumer's churn. Data used in the research to support the findings of the study will to be available by request.

The constructed models are technically correct and effective according to the previously defined criteria for evaluating the success of data research, according to which the quality of the model should be higher than 90%. High level of customer loyalty classification makes it possible to correctly identify the level of customer loyalty and intention to leave the company with an accuracy of over 95%. Since the research results are planned to be used to respond quickly and prevent the client from churn, it is advisable to rank models based on ease of interpretation and accuracy.

As a result of the research, customers were correctly classified according to the potential churn from the network, which will allow in the future to develop the most optimal tariff to reduce the outflow of customers, increase loyalty and attract new clients.

The main factors influencing the outflow of customers from the network, which include average duration of calls and number of calls to other mobile operators, are identified. These factors form a basis for ways of improvement of company's services and growth of business results.

Since the study constructed models showed a high accuracy of customer loyalty classification, i.e., the customer's intention to stay with the company or move to another mobile operator, it is possible to timely apply proactive management to prevent the outflow of customers (because attracting a new customer in 5-10 times more expensive than retaining the existing client), maintaining customer loyalty and satisfaction and reduce costs.

The results of the study can be used to optimize marketing activities of consumer churn management for both Ukrainian and international companies in the telecommunications industry by making effective data-driven decisions and to improve mathematical methodology of consumer churn prediction. So, the main theoretical and practical implications from the research are development of an effective predictive tool for managers to control churn risks and the enriching the literature regarding data analytics and Data Science models for identification of critical factors for customer churn behavior.

In addition, periodic monitoring is necessary to identify new potential factors in a timely manner, which will lead to an outflow of customers. Therefore, it is advisable to conduct research on a regular basis and with additional variables. The limitations of the current research, which cover only 11 variables of consumer's behavior, can be overcome by including additional indicators of customer's characteristics, as well as variables of company's image characteristics (technological, price perception, service perception, etc.), data on consumer's preferences, and so on, that are areas and directions for future research. Models may become obsolete due to a radical change in tariff plans or market conditions, deterioration in the quality of service and communication, a significant change in market position or a significant change in the economic situation. To prevent this situation, it is necessary to improve the models monthly and respond in a timely manner to possible threats.

REFERENCES

Abarca Sánchez, Y., Barreto Rivera, U., Barreto Jara, O., Díaz Ugarte, J.L. (2022). Customer loyalty and retention at a leading telecommunications company in Perú. *Revista Venezolana De Gerencia*, 27(98), 729-743.
https://doi.org/10.52080/rvgluz.27.98.22

Abd-Elrahman, A.H., Ahmed Kamal, J.M. (2022). Relational capital, service quality and organizational performance in the Egyptian telecommunication sector. *International Journal of Emerging Markets*, 17(1), 299-324.
https://doi.org/10.1108/IJOEM-11-2019-0983

Agafonova, A.N., Novikova, E.N., Shakirov, R.A. (2022). New marketing aspects in the digital economy. In: Ashmarina, S.I., Mantulenko, V.V. (eds) Digital Technologies in the New Socio-Economic Reality. ISCDTE 2021. *Lecture Notes in Networks and Systems*, 304, Springer, Cham.
https://doi.org/10.1007/978-3-030-83175-2_65

Aljanabi, A.R.A. (2022). The role of innovation capability in the relationship between marketing capability and new product development: Evidence from the telecommunication sector. *European Journal of Innovation Management*, 25(1), 73-94.
https://doi.org/10.1108/EJIM-04-2020-0146

Al-Shatnwai, A. M., Faris, M. (2020). Predicting customer retention using XGBoost and balancing methods. *International Journal of Advanced Computer Science and Applications*, 11(7), 704-712.
https://doi.org/10.14569/IJACSA.2020.0110785

Bandam, A., Busari, E., Syranidou, C., Linssen, J., Stolten, D. (2022). Classification of building types in Germany: a data-driven modeling approach. *Data*, 7(4), 45.
https://doi.org/10.3390/data7040045

Belbahri, M., Murua, A., Gandouet, O., Nia, V.P. (2021). Qini-based uplift regression. *Annals of Applied Statistics*, 15(3), 1247-1272.
https://doi.org/10.1214/21-AOAS1465

Cacciarelli, D., Boresta, M. (2022). What drives a donor? A machine learning-based approach for predicting responses of nonprofit direct marketing campaigns. *Journal of Philanthropy and Marketing*, 27(2), 1724.
https://doi.org/10.1002/nvsm.1724

Cambier, A., Chardy, M., Figueiredo, R., Ouorou, A., Poss, M. (2022). Optimizing subscriber migrations for a telecommunication operator in uncertain context. *European Journal of Operational Research*, 298(1), 308-321.
https://doi.org/10.1016/j.ejor.2021.06.032

Chernyak, O., Fareniuk, Y. (2020). Modeling of effectiveness of media investment based on Data Science technologies for Ukrainian Bank. *CEUR Workshop Proceedings*, 2732, 282-289.
http://ceur-ws.org/Vol-2732/20200282.pdf

Dadfarnia, M., Matinpour, A.A., Abdoos, M. (2020). Churn prediction in payment terminals using RFM model and deep neural network. 11th International Conference on Information and Knowledge Technology, 98-101.
https://doi.org/10.1109/IKT51791.2020.9345626

De Caigny, A., Coussement, K., Verbeke, W., Idbenjra, K., Phan, M. (2021). Uplift modeling and its implications for B2B customer churn prediction: A segmentation-based modeling approach. *Industrial Marketing Management*, 99, 28-39.
https://doi.org/10.1016/j.indmarman.2021.10.001

De, S., Prabu, P., Paulose, J. (2021). Effective ML techniques to predict customer churn. Proceedings of the 3rd International Conference on Inventive Research in Computing Applications, *ICIRCA*, 895-902.
https://doi.org/10.1109/ICIRCA51532.2021.9544785

Deng, Y., Li, D., Yang, L., Tang, J., Zhao, J. (2021). Analysis and prediction of bank user churn based on ensemble learning algorithm. Proceedings of 2021 IEEE International Conference on Power Electronics, Computer Applications, *ICPECA*, 288-291.
https://doi.org/10.1109/ICPECA51329.2021.9362520

Ding, Y. (2022). Retention strategy for existing users of mobile communications. In: J. Jansen, B., Liang, H., Ye, J. (eds) International Conference on Cognitive based Information Processing and Applications (CIPA 2021). *Lecture Notes on Data Engineering and Communications Technologies*, 84. Springer, Singapore.
https://doi.org/10.1007/978-981-16-5857-0_39

Du, L., Chen, H., Fang, Y., Liang, X., Zhang, Y., Qiao, Y., Guo, Z. (2022). Research on the method of acquiring customer individual demand based on the Quantitative Kano Model. *Comput Intell Neurosci*, 5052711.
https://doi.org/10.1155/2022/5052711

Fang, X. (2021). Research on digital marketing strategy of telecommunication service based on computer complex network model. *Journal of Physics: Conference Series*, 1992(4), 042002.
https://doi.org/10.1088/1742-6596/1992/4/042002

Fedirko, O., Zatonatska, T., Wolowiec, T., Skowron, S. (2021). Data Science and marketing in e-commerce amid COVID-19 pandemic. *European Research Studies Journal*, 2, 3-16.
https://doi.org/10.35808/ersj/2187

Fridrich, M. (2020). Understanding customer churn prediction research with structural topic models. *Economic Computation and Economic Cybernetics Studies and Research*, 54(4), 301-317.
https://doi.org/10.24818/18423264/54.4.20.19

Gartner Research (2021). The Annual Tech Marketing Report: Insights from Gartner's Benchmarks Survey.
https://www.gartner.com/en/documents/4006589

Gattermann-Itschert, T., Thonemann, U.W. (2021). How training on multiple time slices improves performance in churn prediction? *European Journal of Operational Research*, 295(2), 664-674.
https://doi.org/10.1016/j.ejor.2021.05.035

Gopal, P., MohdNawi, N.B. (2021). A survey on customer churn prediction using machine learning and data mining techniques in E-commerce. 2021 IEEE Asia-Pacific Conference on Computer Science and Data Engineering, *CSDE*.
https://doi.org/10.1109/CSDE53843.2021.9718460

Goy, G., Kolukisa, B., Bahcevan, C., Gungor, V.C. (2020). Ensemble churn prediction for internet service provider with machine learning techniques. 5th International Conference on Computer Science and Engineering, *UBMK*, 248-253.
https://doi.org/10.1109/UBMK50275.2020.9219369

Grandhi, B., Patwa, N., Saleem, K. (2021). Data-driven marketing for growth and profitability. *EuroMed Journal of Business*, 16(4), 381-398.
https://doi.org/10.1108/EMJB-09-2018-0054

Gu, J. (2022). Research on precision marketing strategy and personalized recommendation method based on big data drive. Wireless Communications and Mobile Computing.
https://doi.org/10.1155/2022/6751413

Günesen, S.N., Şen, N., Yıldırım, N., Kaya, T. (2021). Customer churn prediction in FMCG sector using machine learning applications, 82-103.
https://doi.org/10.1007/978-3-030-80847-1_6

Havrylovych, M., Kuznietsova, N. (2019). Survival analysis methods for churn prevention in telecommunications industry. *CEUR Workshop Proceedings*, 2577, 47-58.
http://star.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-2577/paper5.pdf

Hemalatha, M., Mahalakshmi, S. (2020). Customer churns prediction in telecom using adaptive logitboost learning approach. *International Journal of Scientific and Technology Research*, 9(2), 5703-5713.
http://www.ijstr.org/final-print/feb2020/Customer-Churns-Prediction-In-Telecom-Using-Adaptive-Logitboost-Learning-Approach.pdf

Hu, D., Zhou, K., Li, F., Ma, D. (2022). Electric vehicle user classification and value discovery based on charging big data. *Energy*, 249, 123698.
https://doi.org/10.1016/j.energy.2022.123698

Huang, J. (2022). Real-time statistical method for marketing profit of Japanese cosmetics online cross-border e-commerce platform. In: Jiang, D., Song, H. (eds) Simulation Tools and Techniques. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, 424. Springer, Cham.
https://doi.org/10.1007/978-3-030-97124-3_48

Jamjoom, A.A. (2021). The use of knowledge extraction in predicting customer churn in B2B. *Journal of Big Data*, 8(110).
https://doi.org/10.1186/s40537-021-00500-3

Jayadi, R., Kelvin, A., Jery, Rifyansyah, P., Mufarih, M., Firmantyo, H.M. (2020). Predicting customer churn of fire insurance policy: a case study in an Indonesian insurance company. Proceedings of the 6th International Conference on Science and Technology, ICST.
https://doi.org/10.1109/ICST50505.2020.9732797

Kelley, K., Todd, M., Hopfer, H., Centinari, M. (2022). Identifying wine consumers interested in environmentally sustainable production practices. *International Journal of Wine Business Research*, 34(1), 86-111.
https://doi.org/10.1108/IJWBR-01-2021-0003

Khrustalova, V., Kononenko, E. (2019). Market of mobile communication services of Ukraine: trends and prospects of development. *Investytsiyi: praktyka ta dosvid*, 1, 37-41.
https://doi.org/10.32702/2306-6814.2019.1.37

Kiguchi, M., Saeed, W., Medi, I. (2022). Churn prediction in digital game-based learning using data mining techniques: logistic regression, decision tree, and random forest. *Applied Soft Computing*, 118.
https://doi.org/10.1016/j.asoc.2022.108491

Kolli, N., Balakrishnan, N. (2020). Hybrid features for churn prediction in mobile telecom networks with data constraints. Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM, 734-741.
https://doi.org/10.1109/ASONAM49781.2020.9381482

Kolomiiets, A., Mezentseva, O., Kolesnikova, K. (2021). Customer churn prediction in the software by subscription models its business using machine learning methods. *CEUR Workshop Proceedings*, 3039, 119-128.
http://star.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-3039/paper49.pdf

Kumar, H., Yadav, R.K. (2020). Rule-based customer churn prediction model using artificial neural network based and Rough Set theory. In: Pant, M., Sharma, T., Verma, O., Singla, R., Sikander, A. (eds) Soft Computing: Theories and Applications. Advances in Intelligent Systems and Computing, 1053. Springer, Singapore.
https://doi.org/10.1007/978-981-15-0751-9_9

Kuznietsova, N., Bidyuk, P. (2018). Forecasting of financial risk users' outflow. IEEE 1st International Conference on System Analysis and Intelligent Computing, SAIC – Proceedings. https://doi.org/10.1109/SAIC.2018.8516782

Kuznietsova, N., Bidyuk, P., Kuznietsova, M. (2022). Data mining methods, models and solutions for Big Data cases in telecommunication industry. In: Babichev, S., Lytvynenko, V. (eds) Lecture Notes in Computational Intelligence and Decision Making. ISDMCI 2021. Lecture Notes on Data Engineering and Communications Technologies, 77. Springer, Cham. https://doi.org/10.1007/978-3-030-82014-5_8

Kuznietsova, N.V. (2017). Information technologies for clients' database analysis and behavior forecasting. *CEUR Workshop Proceedings*, 2067, 56-62. http://star.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-2067

Li, W. (2022). Big Data precision marketing approach under IoT cloud platform information mining. *Comput Intell Neurosci*, 4828108. https://doi.org/10.1155/2022/4828108

Lv, S. (2022). Real estate marketing adaptive decision-making algorithm based on big data analysis. *Security and Communication Networks*, 4(12), 1-11. https://doi.org/10.1155/2022/3443182

Mašić, B., Nešić, S., Vladušić, L. (2018). Challenges in creating transformative growth for companies in digital economy. *ECONOMICS – Innovative and Economics Research Journal*, 6(2), 37-48. https://doi.org/10.2478/eoik-2018-0024

Mo, L., Yang, L. (2022). Research on application effective evaluation of artificial intelligence technology in marketing communication. *Security and Communication Networks*, 3(31), 1-8. https://doi.org/10.1155/2022/3507353

Mykhalchuk, T., Zatonatska, T., Dluhopolskyi, O., Zhukovska, A., Dluhopolska, T., Liakhovych, L. (2021). Development of recommendation system in e-commerce using emotional analysis and machine learning methods. The 11th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS). Vol.1. Cracow, Poland, 527-535. https://ieeexplore.ieee.org/document/9660854

Park, S., Kim, M., Kim, Y., Park, Y. (2022). A deep learning approach to analyze airline customer propensities: the case of South Korea. *Applied Sciences* (Switzerland), 12(4), 1916. https://doi.org/10.3390/app12041916

Priyanga, P., Nadira Banu Kamal, A.R. (2022). Mobile app usage pattern prediction using hierarchical flexi-ensemble clustering (HFEC) for mobile service rating. *Wireless Personal Communications*, 122(4), 3247-3268. https://doi.org/10.1007/s11277-021-09048-0

Qu, Y. (2022). Using data mining techniques to discover customer behavioral patterns for direct marketing. *7th International Conference on Big Data Analytics, ICBDA*, 361-365. https://doi.org/10.1109/ICBDA55095.2022.9760309

Rabiul Alam, M.G., Hussain, S., Mim, M.M.I., Islam, M.T. (2021). Telecom customer behavior analysis using naïve bayes classifier. IEEE 4th International Conference on Computer and Communication Engineering Technology, CCET, 308-312. https://doi.org/10.1109/CCET52649.2021.9544169

Radukic, S., Mastilo, Z., Kostic, Z., Vladusic, L. (2019). Measuring of the goods and labor markets efficiency: comparative study of Western Balkan countries. *Montenegrin Journal of Economics*, 15(2), 95-109. https://doi.org/10.14254/1800-5845/2019.15-2.8

Rakhmatullina, A.R., Shatalova, T.N., Chebykina, M.V. (2022). Conceptual organizational aspects of innovation management processes for industrial enterprises. In: Ashmarina, S.I., Mantulenko, V.V. (eds) Proceedings of the International Conference Engineering Innovations and Sustainable Development. *Lecture Notas in Civic Engineering*. Springer. https://doi.org/10.1007/978-3-030-90843-0_17

Ram, J., Zhang, Z. (2022). Examining the needs to adopt big data analytics in B2B organizations: Development of propositions and model of needs. *Journal of Business and Industrial Marketing*, 37(4), 790-809. https://doi.org/10.1108/JBIM-10-2020-0464

Rezaeian, O., Haghighi, S.S., Shahrabi, J. (2021). Customer churn prediction using data mining techniques for an Iranian payment application. 12th International Conference on Information and Knowledge Technology, IKT, 134-138. https://doi.org/10.1109/IKT54664.2021.9685502

Saanchay, P.M., Thomas, K.T. (2022). An approach for credit card churn prediction using gradient descent. In: IOT with Smart Systems – Smart Innovation, Systems and Technologies, 689-697. Springer Nature Singapore. https://doi.org/10.1007/978-981-16-3945-6_68

Sánchez, D.M., Moreno, A., López, M.D.J. (2022). Machine learning methods for automatic gender detection. *International Journal on Artificial Intelligence Tools*, 31(3). https://doi.org/10.1142/S0218213022410020

Sun Y, Tan X. (2022). Customer relationship management based on SPRINT classification algorithm under Data Mining technology. *Comput Intell Neurosci*, 6170335. https://doi.org/10.1155/2022/6170335

Syaglova, Y.V., Bozhenko, E.S., Larkina, N.G., Polyakova, E.Y., Stefanova, I.V. (2022). Value orientation for marketing customer experience management in companies in a digital transformation. In: Trifonov, P.V., Charaeva, M.V. (eds) Strategies and Trends in Organizational and Project Management. *Lecture Notes in Networks and Systems*, 380. Springer, Cham. https://doi.org/10.1007/978-3-030-94245-8_57

Thakkar, H.K., Desai, A., Ghosh, S., Singh, P., Sharma, G. (2022). Clairvoyant: AdaBoost with cost-enabled cost-sensitive classifier for customer churn prediction. *Comput Intell Neurosci*, 9028580. https://doi.org/10.1155/2022/9028580

Tianyuan, Z., Moro, S. (2021). Research trends in customer churn prediction: a data mining approach. In: Rocha, Á., Adeli, H., Dzemyda, G., Moreira, F., Correia, A.M.R. *World Conference on Information Systems and Technologies*, WorldCIST. Springer. https://doi.org/10.1007/978-3-030-72657-7_22

Todevski, D., Georgieva Svrtinov, V. (2021). Machine learning model for customer churn. KNOWLEDGE – *International Journal*, 47(5), 887-892. https://ikm.mk/ojs/index.php/kij/article/view/4870

Vezzoli, M., Zogmaister, C., Van den Poel, D. (2020). Will they stay or will they go? predicting customer churn in the energy sector. *Applied Marketing Analytics*, 6(2), 136-150. https://www.ingentaconnect.com/content/hsp/ama/2020/00000006/00000002/art00006

Wassouf, W.N., Alkhatib, R., Salloum, K., Balloul, S. (2020). Predictive analytics using big data for increased customer loyalty: Syriatel telecom company case study. *Journal of Big Data*, 7(29). https://doi.org/10.1186/s40537-020-00290-0

Wu, Z., Li, Z. (2021). Customer churn prediction for commercial banks using customer-value-weighted machine learning models. *Journal of Credit Risk*, 17(4), 15-42. https://doi.org/10.21314/JCR.2021.011

Xiahou, X., Harada, Y. (2022). B2C E-commerce customer churn prediction based on K-means and SVM. *Journal of Theoretical and Applied Electronic Commerce Research*, 17(2), 458-475. https://doi.org/10.3390/jtaer17020024

Zatonatska, T., Dluhopolskyi, O., Artyukh, T., Tymchenko, K. (2022). Forecasting the behavior of target segments to activate advertising tools: case of mobile operator Vodafone Ukraine. *ECONOMICS – Innovative and Economics Research Journal, 10(1), 87-104.* https://doi.org/10.2478/eoik-2022-0005

Zatonatska, T., Fedirko, O., Dluhopolskyi, O., Londar, S. (2021). The impact of e-commerce on the sustainable development: case of Ukraine, Poland, and Austria. IOP Conference Series: Earth and Environmental Science, 915 (October 15-16, 2021). Odesa, Ukraine. https://doi.org/10.1088/1755-1315/915/1/012023

Zhang, S., Liao, P., Ye, H., Zhou, Z. (2022). Dynamic marketing resource allocation with two-stage decisions. *Journal of Theoretical and Applied Electronic Commerce Research*, 17(1), 327-344. https://doi.org/10.3390/jtaer17010017

Zhang, T., Moro, S., Ramos, R.F. (2022). A data-driven approach to improve customer churn prediction based on telecom customer segmentation. *Future Internet*, 14(3), 94. https://doi.org/10.3390/fi14030094

Zhu, B., Qian, C., Pan, X., Chen, H. (2020). A trajectory-based deep sequential method for customer churn prediction. *ACM International Conference Proceeding Series*, 114-118. https://doi.org/10.1145/3409073.3409083